Systems
Optimization
Laboratory

Computing Modified Newton Directions Using A Partial Cholesky Factorization

by
Anders Forsgren, Philip E. Gill and Walter Murray

TECHNICAL REPORT SOL-93-1

March 1993

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

Computing Modified Newton Directions Using A Partial Cholesky Factorization

by
Anders Forsgren, Philip E. Gill and Walter Murray

TECHNICAL REPORT SOL-93-1

March 1993

Research and reproduction of this report were partially supported by the National Science Foundation Grants DDM-9204208, DDM-9204547; the Department of Energy Grant DE-FG03-92ER25117 and the Office of Naval Research Grant N00014-90-J-1242.

Research partially supported by the Göran Gustafsson Foundation and the Swedish National Board for Technical Development.

This paper is simultaneously issued as Report TRITA-MAT-1993-9, Department of Mathematics, Royal Institute of Technology; Report LMS 93-2, Department of Mathematics, University of California at San Diego; and Report SOL 93-1, Department of Operations Research, Stanford University. It supersedes part of Report SOL 89-12 *"A modified Newton method for unconstrained minimization"*, Department of Operations Research, Stanford University, 1989.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do NOT necessarily reflect the views of the above sponsors.

Also issued as Operations Research Department Technical Report 93-2. Reproduction in whole or in part is permitted for any purposes of the United States Government. This document has been approved for public release and sale; its distribution is unlimited.

COMPUTING MODIFIED NEWTON DIRECTIONS USING A PARTIAL CHOLESKY FACTORIZATION

Anders FORSGREN*, Philip E. GILL† and Walter MURRAY‡

*Optimization and Systems Theory, Department of Mathematics
Royal Institute of Technology, S-100 44 Stockholm, Sweden

†Department of Mathematics
University of California at San Diego, La Jolla, California 92093-0112, USA

‡Systems Optimization Laboratory, Department of Operations Research
Stanford University, Stanford, California 94305-4022, USA

Technical Report SOL 93-1 §
March 1993

Abstract

The *effectiveness* of Newton's method for finding an unconstrained minimizer of a strictly convex twice continuously differentiable function has prompted the proposal of various *modified* Newton methods for the nonconvex case.

Linesearch modified Newton methods utilize a linear combination of a descent direction and a direction of negative curvature. If these directions are *sufficient* in a certain sense, and a suitable linesearch is used, the resulting method will generate limit points that satisfy the second-order necessary conditions for optimality.

We propose an efficient method for computing a descent direction and a direction of negative curvature that is based on a partial Cholesky factorization of the Hessian. This factorization not only gives theoretically satisfactory directions, but also requires only a *partial* pivoting strategy, i.e., the equivalent of only two rows of the Schur complement need be examined at each step.

Keywords: Unconstrained minimization, modified Newton method, descent direction, negative curvature, Cholesky factorization

---

*Research partially supported by the Göran Gustafsson Foundation and the Swedish National Board for Technical Development.

†‡ Research supported by the Department of Energy Contract DE-FG03-92ER25117, the National Science Foundation Grants DDM-9204208, DDM-9204547, and the Office of Naval Research Grant N00014-90-J-1242.

§This paper is simultaneously issued as Report TRITA-MAT-1993-9, Department of Mathematics, Royal Institute of Technology; Report LMS 93-2, Department of Mathematics, University of California at San Diego; and Report SOL 93-1, Department of Operations Research, Stanford University. It supersedes part of Report SOL 89-12 *"A modified Newton method for unconstrained minimization"*, Department of Operations Research, Stanford University, 1989.

1. Introduction

We consider the unconstrained minimization of a twice continuously differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. If f is strictly convex, the excellent local convergence properties of Newton's method make it one of the most effective methods for minimization (see, e.g., Ortega and Rheinboldt [OR70]).

In the non-convex case, various *modified* Newton methods have been proposed that ensure convergence from an arbitrary starting point. Here we focus on the class of *linesearch* modified Newton methods (for a complete discussion of modified Newton methods and their relative merits, see, e.g., Shultz *et al.* [SSB85], Dennis and Schnabel [DS89]). Linesearch modified Newton methods generate a sequence $\{x_k\}_{k=0}^{\infty}$ of improving estimates of a local minimizer. At iteration k, a linesearch is performed along a path formed from a linear combination of two directions s_k and d_k, where either s_k or d_k can be zero. The directions s_k and d_k are chosen such that $g_k^T s_k \leq 0$ and $d_k^T H_k d_k \leq 0$, where g_k and H_k denote the gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ evaluated at x_k. (Implicitly, we also assume the condition $g_k^T d_k \leq 0$, which can be imposed with a trivial sign change of d_k.) Each nonzero s_k satisfies $g_k^T s_k < 0$ and is known as a *descent direction*. Each nonzero d_k satisfies $d_k^T H_k d_k < 0$ and is known as a *direction of negative curvature*. If d_k is nonzero, H_k must have at least one negative eigenvalue. (Henceforth we will sacrifice precision for the sake of brevity and refer to the sequences $\{s_k\}$ and $\{d_k\}$ as sequences of "descent directions" and "directions of negative curvature".) Linesearch methods of this type have been proposed by Gill and Murray [GM74], Fletcher and Freeman [FF77], McCormick [McC77], Mukai and Polak [MP78], Kaniel and Dax [KD79], and Goldfarb [Gol80].

Moré and Sorensen [MS79] have shown that if: (i) a modified Newton method is used in conjunction with a suitable linesearch; and (ii) the directions s_k and d_k are *sufficient* in the sense that the sequences $\{s_k\}$ and $\{d_k\}$ are bounded and satisfy

$$g_k^T s_k \rightarrow 0 \quad \Rightarrow \quad g_k \rightarrow 0 \quad \text{and} \quad s_k \rightarrow 0, \tag{1.1a}$$

and

$$d_k^T H_k d_k \rightarrow 0 \quad \Rightarrow \quad \min\{\lambda_{\min}(H_k), 0\} \rightarrow 0 \quad \text{and} \quad d_k \rightarrow 0, \tag{1.1b}$$

then every limit point of the resulting sequence $\{x_k\}_{k=0}^{\infty}$ will satisfy the second-order necessary conditions for optimality.

It has been observed in practice that the number of iterates at which the Hessian is positive definite is large compared to the total number of iterations. Since linesearch methods revert to Newton's method when the Hessian is sufficiently positive definite, it would seem sensible to use a modified Newton method based on the most efficient method for solving a symmetric positive-definite system. This is the motivation for the modified Cholesky factorization proposed by Gill and Murray [GM74]. However, it has been shown by Moré and Sorensen [MS79] that this factorization may not give directions of negative curvature that are sufficient in the sense of (1.1b). This paper is motivated by the need for an algorithm with the efficiency and simplicity of the Cholesky factorization, but with the guarantee of convergence when used with a suitable linesearch. It is shown in Section 3 that a

partial Cholesky factorization can give search directions that are sufficient in the sense of (1.1).

To simplify the notation, we will drop the subscript k when referring to the quantities g_k, H_k, s_k and d_k at a specific iteration. Unless otherwise stated, $\|\cdot\|$ refers to the vector two-norm or its induced matrix norm. The vector e_j denotes the j-th unit vector whose dimension is determined by the context.

2. The partial Cholesky factorization

The partial Cholesky factorization of H is a variant of the standard Cholesky factorization with diagonal pivoting. The algorithm is stated in *outer-product form*, where the Schur complement associated with the unfactorized part of H is updated explicitly at each step (see, e.g., Golub and Van Loan [GV89, page 143] and Higham [Hig90]).

At each step, the largest diagonal is selected as pivot and is used to eliminate a row and column from the Schur complement. The algorithm continues until either all the matrix has been factorized or the pivot is considered unacceptable. The final factors are therefore uniquely determined by the rule used to accept the pivot (i.e., the rule used to terminate the elimination). Termination is controlled by a preassigned scalar parameter ν ($0 < \nu < 1$). A pivot is acceptable if it is both positive and larger in absolute value than ν times the off-diagonal of largest magnitude in the pivot row and column. At each step, the determination of an acceptable pivot requires the examination of the diagonals and a single row of the Schur complement. (For a similar scheme in the context of quadratic programming, see Casas and Pola [CP90].)

It will be shown below that once a pivot is deemed unacceptable (and hence the factorization is terminated), a suitable direction of negative curvature can be determined from the elements of the remaining Schur complement.

Let P denote the permutation matrix representing the symmetric interchanges performed during the factorization. If n_1 denotes the number of steps needed before termination, the factorization implicitly identifies a leading $n_1 \times n_1$ positive-definite submatrix of the permuted matrix $P^T H P$. In terms of a partition H_{11}, H_{12}, H_{21} and H_{22} of $P^T H P$, we have

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & \\ L_{21} & I \end{pmatrix} \begin{pmatrix} B_1 & \\ & B_2 \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ & I \end{pmatrix}, \qquad (2.1)$$

where L_{11} is unit lower triangular and B_1 is a positive-definite diagonal matrix. The submatrix H_{11} is positive definite, and $H_{11} = L_{11} B_1 L_{11}^T$ is its usual Cholesky factorization obtained using diagonal pivoting. The factorization may be written briefly as $H = L B L^T$, where L is a row-permuted lower-triangular matrix with

$$L = P \begin{pmatrix} L_{11} & \\ L_{21} & I \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} B_1 & \\ & B_2 \end{pmatrix}. \qquad (2.2)$$

We will use n_2 to denote the size of H_{22}, so that $n_1 + n_2 = n$. A "pseudo-matlab" version of the partial Cholesky algorithm is given in Algorithm 2.1.

The curvature along any direction d computed from the partial Cholesky factorization is related to the magnitude of the smallest eigenvalue of the Schur complement B_2. The following lemma relates the smallest eigenvalue of B_2 to the smallest eigenvalue of H.

Lemma 2.1. *Let H be a symmetric $n \times n$ matrix with at least one negative eigenvalue. Let the partial Cholesky factorization of H be denoted by $H = LBL^T$, where P^THP is partitioned as in (2.1). Then*

$$\lambda_{\min}(B_2) \leq \lambda_{\min}(H) \quad \textit{and} \quad B_2 = Y^THY,$$

where

$$Y = P \begin{pmatrix} -L_{11}^{-T}L_{21}^T \\ I \end{pmatrix} = P \begin{pmatrix} -H_{11}^{-1}H_{12} \\ I \end{pmatrix}. \tag{2.3}$$

Proof. The inequality $\lambda_{\min}(B_2) \leq \lambda_{\min}(H)$ can be established using the identity

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} 0 & \\ & B_2 \end{pmatrix} + \begin{pmatrix} L_{11} \\ L_{21} \end{pmatrix} B_1 \begin{pmatrix} L_{11}^T & L_{21}^T \end{pmatrix}, \tag{2.4}$$

which is a rearrangement of the factorization (2.1). The eigenvalues of H and P^THP are identical. Moreover, the positive-definiteness of B_1 implies that the second term on the right-hand-side of (2.4) is positive semidefinite. Since the eigenvalues of P^THP cannot increase on subtraction of a positive semidefinite matrix, it must follow that $\min\{0, \lambda_{\min}(B_2)\} \leq \lambda_{\min}(H)$ (see e.g., Golub and Van Loan [GV89, page 411]). From the assumption $\lambda_{\min}(H) < 0$, we conclude that $\lambda_{\min}(B_2) \leq \lambda_{\min}(H)$, as required.

To show that the matrix Y (2.3) is well defined, it is sufficient to verify that $H_{11}^{-1}H_{12} = L_{11}^{-T}L_{21}^T$. This is an immediate consequence of multiplying the partitioned right-hand-side matrix from (2.1) to obtain $H_{11} = L_{11}B_1L_{11}^T$ and $H_{12} = L_{11}B_1L_{21}^T$.

Finally, the identity $Y^THY = B_2$ may be verified by expressing $L^{-1}HL^{-T} = B$ in the partitioned form

$$\begin{pmatrix} L_{11}^{-1} & \\ -L_{21}L_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} L_{11}^{-T} & -L_{11}^{-T}L_{21}^T \\ & I \end{pmatrix} = \begin{pmatrix} B_1 & \\ & B_2 \end{pmatrix},$$

from which the result follows. ∎

Note that the matrix Y (2.3) consists of the last n_2 columns of L^{-T}. Our analysis requires bounds on the norms of Y, L and L^{-1}, which are provided by the following lemma given by Higham [Hig90].

Lemma 2.2. *Let H be factorized using the partial Cholesky factorization described in Algorithm 2.1. If P^THP is partitioned as in (2.1), then*

(a) $\|L_{11}^{-T}L_{21}^T\| \leq \dfrac{1}{\nu}\sqrt{\dfrac{1}{3}(n-n_1)(4^{n_1}-1)}$;

(b) $\|\bar{L}_{11}^{-T} L_{21}^T e_i\| \leq \frac{1}{\nu}\sqrt{\frac{1}{3}(4^{n_1}-1)}$;

(c) $\|L\| \leq \frac{n}{\nu}$;

(d) $\|L^{-1}\| \leq \frac{n}{\nu} 2^{n_1-1}$.

Proof. Part (a) follows immediately from Lemma 9.4 of Higham [Hig90] and the fact that the elements of L_{21} are bounded in absolute value by $1/\nu$. Part (b) is a consequence of part (a), since $L_{21}^T e_i$ is an n_2-vector whose elements are bounded in absolute value by $1/\nu$. Part (c) follows from the fact that all elements of L are bounded by $1/\nu$ in absolute value. Similarly, part (d) is a consequence of the fact that all elements of L^{-1} are bounded by $2^{n_1-1}/\nu$ (see Higham [Hig90] for details). ∎

2.1. Computation of the descent direction

We now discuss the application of the partial Cholesky factorization to the calculation of a descent direction s_k satisfying (1.1a). Let $\bar{B}$ be any positive-definite modification of B, i.e., $\bar{B}$ is a positive-definite matrix with $\|B - \bar{B}\|$ "small" and $\bar{B} = B$ when B is sufficiently positive definite. There are many choices for $\bar{B}$—for example, consider the block-diagonal matrix $\bar{B} = \mathrm{diag}(B_1, I)$, where I is the identity matrix of order n_2. With this definition, when $n_1 = n$ and H is sufficiently positive definite, $\bar{B}_1 = B_1$ and s satisfies the usual Newton equations $Hs = -g$.

Lemma 2.3. *Let H be factorized using the partial Cholesky factorization described in Algorithm 2.1 and assume that $P^T H P$ is partitioned as in (2.1). Let $\bar{B}$ be a positive-definite modification of B, and let s satisfy*

$$L\bar{B}L^T s = -g. \tag{2.5}$$

Then,

$$-g^T s \geq \frac{\nu^2}{n^2 \lambda_{\max}(\bar{B})}\|g\|^2 \quad \text{and} \quad \|s\| \leq \frac{n^2 4^{n_1-1}}{\nu^2 \lambda_{\min}(\bar{B})}\|g\|.$$

Proof. From the definition of s in (2.5) we have

$$s = -L^{-T}\bar{B}^{-1}L^{-1}g. \tag{2.6}$$

Premultiplying (2.6) by g^T gives

$$-g^T s = g^T L^{-T}\bar{B}^{-1}L^{-1}g \geq \frac{1}{\|L\|^2 \lambda_{\max}(\bar{B})} g^T g,$$

and the required lower bound on $-g^T s$ follows from part (c) of Lemma 2.2. To obtain the bound on $\|s\|$ we derive the inequality $\|s\| \leq \lambda_{\max}(\bar{B}^{-1})\|L^{-1}\|^2\|g\|$, by taking norms of both sides of (2.6), substituting for L from (2.2) and using norm inequalities. The required upper bound follows from part (d) of Lemma 2.2. ∎

2.2. Computation of the direction of negative curvature

The formula for d is derived from a method for computing directions of negative curvature in quadratic programming (see Forsgren *et al.* [FGM91]). The approach is based on the observation that, in the positive-definite case, the Newton direction is a minimizer of a quadratic model with gradient g and Hessian H. In particular, the Newton direction can be found by a quadratic programming algorithm that minimizes the model function while successively releasing variables from temporarily fixed values. This analogy can be extended to the indefinite case, where the variables corresponding to H_{22} are temporarily fixed at their current values, and a direction of negative curvature is defined by releasing either one or two of the fixed variables. This scheme corresponds to using a direction of negative curvature that is a multiple of either y_i or $y_i \pm y_j$, where y_i and y_j denote columns i and j of the matrix Y (2.3). The following lemma shows how the indices i and j are determined from the elements of $B_2 = Y^T H Y$.

Lemma 2.4. *On termination of the partial Cholesky factorization with diagonal pivoting, let $P^T H P$ be partitioned as in (2.1). If $n_1 = n$, define $d = 0$. Otherwise, if $n_1 < n$, define d as follows. Given $\rho = \max_{i>n_1, j>n_1} |b_{ij}|$ and any pair of indices q $(q > n_1)$ and r $(r > n_1)$ such that $|b_{qr}| = \rho$, let d be the solution of*

$$L^T d = \sqrt{\rho}\, v, \quad \text{where} \quad v = \begin{cases} e_q & \text{if } q = r, \\ \dfrac{1}{\sqrt{2}}(e_q - \operatorname{sgn}(b_{qr}) e_r) & \text{otherwise.} \end{cases}$$

Then, if $\lambda_{\min}(H) \geq 0$, then $d = 0$. Otherwise, if $\lambda_{\min}(H) < 0$, then

$$-\frac{1}{n_2}\lambda_{\min}(H) \leq d^T d \leq -\frac{1}{1-\nu}\left(1 + \frac{2(4^{n_1}-1)}{3\nu^2}\right)^2 \lambda_{\min}(H)$$

and

$$\frac{d^T H d}{d^T d} \leq \frac{3\nu^2(1-\nu)}{n_2(3\nu^2 + 2(4^{n_1}-1))}\lambda_{\min}(H).$$

Proof. If $n_1 = n$, then $\lambda_{\min}(H) > 0$, and the lemma holds from the definition $d = 0$. For the remainder of the proof, assume that $n_1 < n$.

First, it is necessary to show that $\gamma \leq \nu\rho$, where $\gamma = \max\{\{\max_{i>n_1} b_{ii}\}, 0\}$. If the factorization terminates with $\gamma = 0$, the inequality $\gamma \leq \nu\rho$ is trivially satisfied. If the factorization terminates with $\gamma > 0$, there exists an index t $(t > n_1)$ such that $b_{tt} = \gamma$. Since γ must be an unacceptable pivot, we can infer that $\gamma < \nu \max_{i \neq t, i>n_1} |b_{it}|$. Consequently, if $n_1 < n$, it must hold that $\gamma \leq \nu\rho$.

Let d_1 and v_1 denote the first n_1 components of Pd and v respectively. Similarly, let d_2 and v_2 denote the last n_2 components of Pd and v. The definitions of d and v imply that $\|v_1\| = 0$, $\|v_2\| = 1$, and $d_2 = \sqrt{\rho}\, v_2$. Therefore,

$$d^T d = d_1^T d_1 + d_2^T d_2 \geq \rho v_2^T v_2 = \rho. \tag{2.7}$$

Similarly, the definition of d and (2.2) imply that

$$d^T d \leq (1 + \|L_{11}^{-T} L_{21}^T v\|^2)\rho \leq \left(1 + \frac{2(4^{n_1} - 1)}{3\nu^2}\right)\rho, \tag{2.8}$$

where the last inequality follows from Lemma 2.2. Combining (2.7) and (2.8) yields

$$\rho \leq d^T d \leq \left(1 + \frac{2(4^{n_1} - 1)}{3\nu^2}\right)\rho. \tag{2.9}$$

Consider the case $\rho = 0$, which is equivalent to H being positive semidefinite and singular with $\lambda_{\min}(H) = 0$. In this case, (2.9) implies $d = 0$, as required.

Now assume that $\rho > 0$. First, if $q = r$, then $|b_{qq}| = \rho$. Since $b_{qq} \leq \gamma \leq \nu\rho < \rho$, it must hold that $b_{qq} = -\rho$, and from the definition of d we obtain the bound

$$d^T H d = \rho b_{qq} \leq -(1 - \nu)\rho^2. \tag{2.10}$$

Alternatively, if $q \neq r$, then the definition of d yields

$$d^T H d = \frac{\rho}{2}(b_{qq} + b_{rr} - 2|b_{qr}|) \leq \rho(\gamma - \rho) \leq -(1 - \nu)\rho^2, \tag{2.11}$$

where the inequalities follow from the conditions $b_{qq} \leq \gamma$, $b_{rr} \leq \gamma$ and $\rho \geq \gamma/\nu$.

Since the magnitude of every element in B_2 is bounded by ρ, the Gershgorin circle theorem and Lemma 2.1 imply

$$\rho \geq -\frac{1}{n_2}\lambda_{\min}(B_2) \geq -\frac{1}{n_2}\lambda_{\min}(H). \tag{2.12}$$

Combining (2.9), (2.10), (2.11) and (2.12) we obtain

$$\frac{d^T H d}{d^T d} \leq -\frac{3\nu^2(1 - \nu)}{3\nu^2 + 2(4^{n_1} - 1)}\rho \leq -\frac{3\nu^2(1 - \nu)}{n_2(3\nu^2 + 2(4^{n_1} - 1))}\lambda_{\min}(H), \tag{2.13}$$

as required.

Since, by definition, $\lambda_{\min}(H) \leq d^T H d / d^T d$, the left-most inequality of (2.13) gives an upper bound on ρ, which in conjunction with (2.9) and (2.12) give the bounds on $d^T d$ as

$$-\frac{1}{n_2}\lambda_{\min}(H) \leq d^T d \leq -\frac{1}{1 - \nu}\left(1 + \frac{2(4^{n_1} - 1)}{3\nu^2}\right)^2 \lambda_{\min}(H).$$

∎

This lemma gives a relation between the curvature along d and the smallest eigenvalue of H, which is the "best possible" curvature. The bound is exponential in n_1, but the computational experiments discussed below imply that the bound is unlikely to be tight in practice. However, as in Higham [Hig90], we observe that

there do exist matrices whose bound is "almost" tight. For given n ($n \geq 3$) and θ, define $L(\theta)$ and $B(\theta)$ as

$$L(\theta) = \begin{pmatrix} 1 & & & & & & \\ -\cos\theta & 1 & & & & & \\ -\cos\theta & -\cos\theta & 1 & & & & \\ \vdots & \vdots & \ddots & \ddots & & & \\ -\cos\theta & -\cos\theta & \cdots & -\cos\theta & 1 & & \\ -\cos\theta & -\cos\theta & \cdots & -\cos\theta & -\cos\theta & 1 & \\ -\cos\theta & -\cos\theta & \cdots & -\cos\theta & -\cos\theta & 0 & 1 \end{pmatrix} \quad \text{and}$$

$$B(\theta) = \begin{pmatrix} 1 & & & & & & \\ & \sin^2\theta & & & & & \\ & & \sin^4\theta & & & & \\ & & & \ddots & & & \\ & & & & \sin^{2(n-3)}\theta & & \\ & & & & & 0 & -1 \\ & & & & & -1 & 0 \end{pmatrix}.$$

Define $H(\theta) = L(\theta)B(\theta)L(\theta)^T$. If $\theta = 0$, it is shown in Lemma A.1 of Appendix A that $\lambda_{\min}(H(0)) = -\frac{1}{2}(\sqrt{n^2 + 2n - 7} - n + 1)$, where

$$-1 \leq \lambda_{\min}(H(0)) \leq -1 + \frac{4}{n+1}.$$

If $\theta = 0$, the partial Cholesky factorization with diagonal pivoting gives $n_1 = 1$. If $d(\theta)$ denotes the direction of negative curvature associated with $H(\theta)$, we obtain

$$\frac{d(0)^T H(0) d(0)}{d(0)^T d(0)} = -\frac{1}{3}, \tag{2.1.1}$$

and $d(0)$ is a satisfactory direction of negative curvature. However, if θ is nonzero, it follows from the analysis of Higham [Hig90] that the partial Cholesky factorization with diagonal pivoting will define $L(\theta)$ and $B(\theta)$ as factors with $n_1 = n - 2$ for all $\theta \neq 0$. Moreover,

$$\lim_{\theta \to 0} \frac{d(\theta)^T H(\theta) d(\theta)}{d(\theta)^T d(\theta)} = -\frac{3}{1 + 2 \cdot 4^{n-2}},$$

and for θ near zero, the curvature along $d(\theta)$ is close to the worst possible value predicted by Lemma 2.4 (see Higham [Hig90] for the details). This "pathological" example arises because the principal submatrix of order $n - 2$ of $H(\theta)$ is positive definite but arbitrarily close to being singular so that $\|H_{11}^{-1} H_{12}\|$ (or equivalently $\|L_{11}^{-T} L_{21}^T\|$) is very large. This is reflected in arbitrarily small pivot elements.

A numerical experiment was devised to investigate if the bound of Lemma 2.4 is likely to be sharp for an arbitrary indefinite matrix. Matlab 4.0 was used to generate directions of negative curvature for a large set of random indefinite symmetric

matrices of order 50. Each H was defined as $Q\Lambda Q^T$, with Q a random orthogonal matrix and Λ a random diagonal matrix with at least one negative element. The matrix Q was obtained from the QR-factorization of a 50×50 matrix whose elements were taken from an independent normal distribution with zero mean and unit variance. The elements of Λ were taken from an independent uniform distribution in the interval $[-25, 25]$. Directions of negative curvature were computed with ν-values $\sqrt{\epsilon}$, 0.05, 0.10, ..., 0.95, and $1 - \sqrt{\epsilon}$, where ϵ denotes the machine precision. A new random matrix was generated for each factorization, giving a total of 1500 matrices for each value of ν. Figure 2.1 gives the outcome of the computational experiment. The three lines depict the maximum, mean, and minimum values of the ratio r of $d^T H d / d^T d$ to $\lambda_{\min}(H)$. Each "+" represents the value of r for a particular value of the parameter ν.

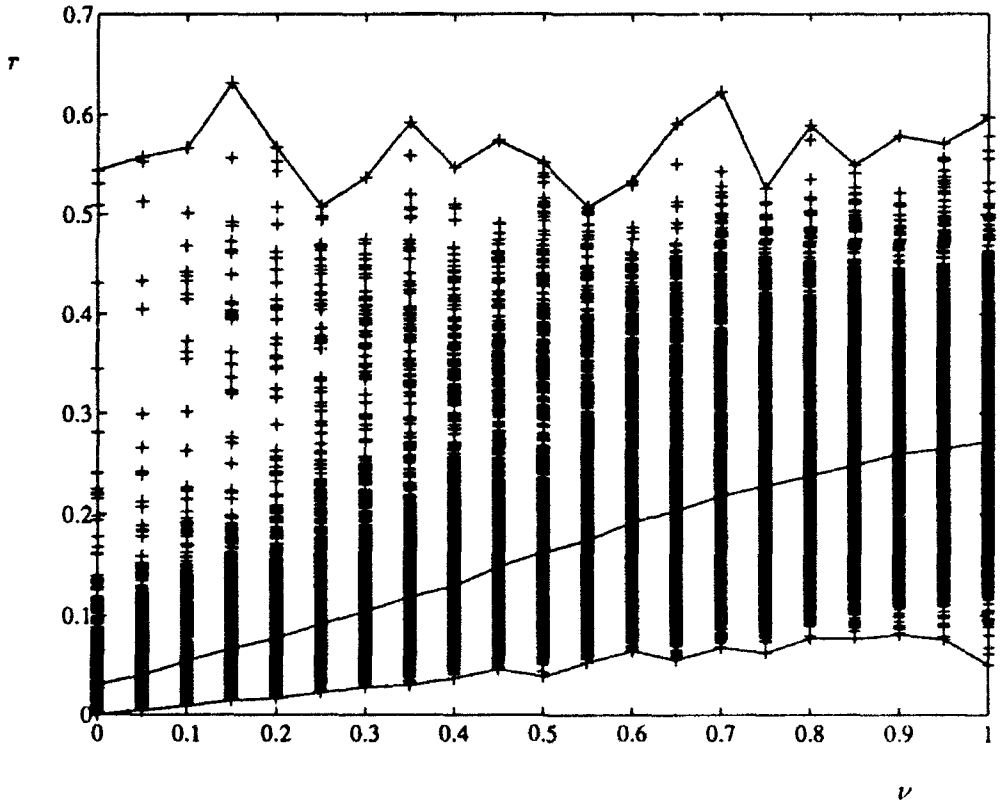

Figure 2.1: Curvature ratio r as function of ν.

The bound on r given by Lemma 2.4 is approximately maximized for $\nu = 2/3$. If, for $n = 50$, this optimal value gives $n_1 = 49$, the theoretical bound is approximately 7×10^{-31}. This should be compared with the computed values of r, which never fell below 0.05 when ν was larger than 0.5. The minimum value of r attained a maximum of 0.0809 for $\nu = 0.9$. Based on these results, we would recommend a

value of ν in the range (0.5, 0.95). Note that the larger the value of ν, the smaller the value of n_1 and consequently, the smaller the amount of computation.

3. Theoretical results

The partial Cholesky factorization can be used as the basis for a descent method for minimizing a twice-continuously differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. This method defines a sequence $\{x_k\}_{k=0}^{\infty}$ of improving estimates of a local minimizer.

Let x_0 be any starting point such that the level set $\{x \mid f(x) \leq f(x_0)\}$ is compact. Let $\{s_k\}$ and $\{d_k\}$ be bounded sequences such that each s_k is a descent direction that satisfies (1.1a) and each d_k is a direction of negative curvature that satisfies (1.1b). Moré and Sorensen [MS79] show that with an appropriate linesearch, certain linear combinations of s_k and d_k define x_{k+1} so that every limit point of $\{x_k\}_{k=0}^{\infty}$ will satisfy the second-order necessary conditions for optimality—i.e., at every limit point $\bar{x}$, $\nabla f(\bar{x})$ is zero and $\nabla^2 f(\bar{x})$ is positive semidefinite. The main result of this paper—that the search directions obtained using the partial Cholesky factorization are sufficient in the sense of Moré and Sorensen [MS79]—is stated in the following theorem.

Theorem 3.1. *Let $\{x_k\}_{k=0}^{\infty}$ be a sequence of iterates contained in a compact region of $\mathbb{R}^n$, and assume that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a twice-continuously differentiable function. For each k, define $g_k = \nabla f(x_k)$ and $H_k = \nabla^2 f(x_k)$, and let $H_k = L_k B_k L_k^T$ be the partial Cholesky factorization of H_k as described in Algorithm 2.1. Given positive constants c_1 and c_2 $(c_1 < c_2)$, let s_k be defined from Lemma 2.3 with the additional requirement that $c_1 \leq \lambda_{\min}(\bar{B}_k) \leq \lambda_{\max}(\bar{B}_k) \leq c_2$. Finally, let d_k be defined from Lemma 2.4. Then, $\{s_k\}$ and $\{d_k\}$ are bounded sequences such that*

$$g_k^T s_k \rightarrow 0 \quad \Rightarrow \quad g_k \rightarrow 0 \quad \textit{and} \quad s_k \rightarrow 0$$

and

$$d_k^T H_k d_k \rightarrow 0 \quad \Rightarrow \quad \min\{\lambda_{\min}(H_k), 0\} \rightarrow 0 \quad \textit{and} \quad d_k \rightarrow 0.$$

Proof. Since $\{x_k\}$ lies in a compact region, the smoothness of f implies that $\{\|g_k\|\}$ and $\{\|H_k\|\}$ are bounded.

With the existence of c_1 and c_2, and the boundedness of $\|g_k\|$, Lemma 2.3 implies that $\{s_k\}$ is a bounded sequence, and $g_k^T s_k \rightarrow 0$ implies $g_k \rightarrow 0$ and $s_k \rightarrow 0$, as required.

Lemma 2.4 and the boundedness of $\|H_k\|$ imply that $\{d_k\}$ is a bounded sequence, and $d_k^T H_k d_k \rightarrow 0$ implies $d_k \rightarrow 0$ and $\min\{\lambda_{\min}(H_k), 0\} \rightarrow 0$, as required. ∎

If $\nabla^2 f(x_k)$ is sufficiently positive definite, all pivots will be acceptable and the partial Cholesky factorization will terminate with $n_1 = n$. This implies that if $\{x_k\}_{k=0}^{\infty}$ has a limit point $\bar{x}$ at which $\nabla^2 f(\bar{x})$ is sufficiently positive definite, then the iterates will be identical to those of Newton's method for k sufficiently large.

4. Discussion

The partial Cholesky factorization may be implemented in other ways. For example, the calculation of the matrix H_{11} can be made independent of the calculation of the descent direction s_k. Once a direction of negative curvature has been defined, a descent direction can be calculated by forming the modified Cholesky factorization of B_2 (see, e.g., Gill and Murray [GM74], Schnabel and Eskow [SE90]).

The algorithm of Section 2.2 requires the examination of the diagonals and a single row of the Schur complement at each step. Alternative strategies can be devised in which the complete Schur complement is examined under certain exceptional circumstances. For example, if a pivot is small, the pivot acceptance criterion could be strengthened so that a pivot is acceptable if, in addition to the requirements of Algorithm 2.1, it is larger in absolute value than $\nu b_{\max}$, where $b_{\max}$ is either the diagonal of largest magnitude in the Schur complement or the element of largest magnitude in the full Schur complement. Each of these modifications gives an algorithm with identical theoretical properties, but a potentially smaller value of n_1. However, this potential improvement is at the expense of an increase in the number of comparisons during the factorization. The pivot criterion that requires the examination of the full Schur complement would cope successfully with the "pathological" $H(\theta)$ of Section 2.2 since the factorization would terminate after one step for θ sufficiently small.

5. Summary

We have shown how a partial Cholesky factorization can be used to define search directions suitable for a linesearch-based modified Newton method. The resulting directions are sufficient in the sense that it is possible to generate a sequence $\{x_k\}_{k=0}^{\infty}$ with limit points having a zero gradient and a positive-semidefinite Hessian.

To our knowledge, this is the first triangular factorization that not only gives theoretically satisfactory directions, but also requires only a *partial* pivoting strategy, i.e., the equivalent of only two rows of the Schur complement need be examined at each step.

Acknowledgement

We thank the Numerical Algorithms Group, Oxford, for providing the computing facilities that enabled work on this paper to be completed.

A. Eigenvalues of $H(0)$

Lemma A.1. *Let the $n \times n$-matrices $L(0)$ and $B(0)$ be defined as in Section 2 for $\theta = 0$ and $n \geq 3$. Define $H(0) = L(0)B(0)L(0)^T$. Then $\lambda = -\frac{1}{2}(\sqrt{n^2 + 2n - 7} - n + 1)$ is the smallest eigenvalue of $H(0)$, and $-1 \leq \lambda \leq -1 + 4/(n+1)$.*

Proof. It is straightforward to verify that

$$H(0) = \begin{pmatrix} 1 & -1 & -1 & \cdots & -1 & -1 & -1 \\ -1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ -1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ -1 & 1 & 1 & \cdots & 1 & 1 & 0 \\ -1 & 1 & 1 & \cdots & 1 & 0 & 1 \end{pmatrix}.$$

Since $B(0)$ has one negative eigenvalue and $L(0)$ is nonsingular, Sylvester's law of inertia implies that $H(0)$ has one negative eigenvalue (see e.g., Golub and Van Loan [GV89, page 416]). Consequently, since λ is negative for $n \geq 3$, it is enough to show that it is an eigenvalue.

Assume that $v = (1 \ \ -1 \ \ -1 \cdots -1 \ \ a \ \ a)^T$ is an eigenvector of $H(0)$ for some scalar a. Then, if v is an eigenvector, there must exist a λ such that

$$n - 2 - 2a = \lambda \quad \text{and} \tag{A.1a}$$

$$-n + 2 + a = \lambda a. \tag{A.1b}$$

It is straightforward to show that for $n \geq 3$, (A.1) has a negative solution λ given by

$$\lambda = -\frac{\sqrt{n^2 + 2n - 7} - n + 1}{2} \quad \text{and} \quad a = \frac{\sqrt{n^2 + 2n - 7} + n - 3}{4}.$$

The upper and lower bounds on λ follow from the sequence of inequalities

$$n + 1 \geq \sqrt{(n+1)^2 - 8} = (n+1)\sqrt{1 - \frac{8}{(n+1)^2}} \geq n + 1 - \frac{8}{n+1}.$$

(Note that the lower bound can also be obtained directly from Lemma 2.1.) ∎

References

[CP90] E. Casas and C. Pola. An algorithm for indefinite quadratic programming based on a partial Cholesky factorization. Preprint. Universidad de Cantabria, Santander, Spain, 1990.

[DS89] J. E. Dennis, Jr. and R. B. Schnabel. A view of unconstrained optimization. In G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors, *Handbooks in Operations Research and Management Science, volume 1. Optimization*, chapter 1, pages 1–72. North Holland, Amsterdam, New York, Oxford and Tokyo, 1989.

[FF77] R. Fletcher and T. L. Freeman. A modified Newton method for minimization. *J. Optimization Theory and Applications*, 23, 357–372, 1977.

[FGM91] A. Forsgren, P. E. Gill, and W. Murray. On the identification of local minimizers in inertia-controlling methods for quadratic programming. *SIAM J. on Matrix Analysis and Applications*, 12, 730–746, 1991.

[GM74] P. E. Gill and W. Murray. Newton-type methods for unconstrained and linearly constrained optimization. *Mathematical Programming*, 7, 311–350, 1974.

[Gol80] D. Goldfarb. Curvilinear path steplength algorithms for minimization which use directions of negative curvature. *Mathematical Programming*, 18, 31–40, 1980.

[GV89] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, second edition, 1989.

[Hig90] N. J. Higham. Analysis of the Cholesky decomposition of a semi-definite matrix. In M. G. Cox and S. Hammarling, editors, *Reliable Numerical Computation*, pages 161–185. Oxford University Press, 1990.

[KD79] S. Kaniel and A. Dax. A modified Newton's method for unconstrained minimization. *SIAM J. on Numerical Analysis*, 16, 324–331, 1979.

[McC77] G. P. McCormick. A modification of Armijo's step-size rule for negative curvature. *Mathematical Programming*, 13, 111–115, 1977.

[MP78] H. Mukai and E. Polak. A second-order method for unconstrained optimization. *J. Optimization Theory and Applications*, 26, 501–513, 1978.

[MS79] J. J. Moré and D. C. Sorensen. On the use of directions of negative curvature in a modified Newton method. *Mathematical Programming*, 16, 1–20, 1979.

[OR70] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[SE90] R. B. Schnabel and E. Eskow. A new modified Cholesky factorization. *SIAM J. on Scientific and Statistical Computing*, 11, 1136–1158, 1990.

[SSB85] G. A. Shultz, R. B. Schnabel, and R. H. Byrd. A family of trust-region based algorithms for unconstrained minimization with strong global convergence properties. *SIAM J. on Numerical Analysis*, 22, 47–67, 1985.

Algorithm 2.1. *An algorithm for the partial Cholesky factorization*

```
%PARTCHOL   Partial Cholesky factorization routine for a real symmetric
%           matrix H.
%           [L,B,perm,n1] = partchol(H)
%           forms a permutation perm, a unit lower-triangular matrix
%           L(perm,:) and a block diagonal matrix B such that L·B·L'=H
%           using the partial Cholesky factorization with diagonal pivoting.
%           The size of the positive-definite principal submatrix obtained
%           in the factorization is denoted by n1.
function [L,B,perm,n1] = partchol(H)
n = length(H);
perm = 1:n;
B = H;
L = zeros(n);
ν ∈ (0,1);
k = 1;
n1 = 0;
while k ≤ n
   [μr,r] = max([zeros(1,k-1) diag(B(k:n,k:n))']);
   if k < n
      μpr = max(abs(B(r,[1:r-1 r+1:n])));
   else
      μpr = 0;
   end
   if μr > 0 and μr ≥ ν · μpr
      n1 = k;
      perm([k r]) = perm([r k]);
      B([k r],:) = B([r k],:);
      B(:,[k r]) = B(:,[r k]);
      L(perm(k:n),k) = B(k:n,k)/B(k,k);
      if k < n
         B(k+1:n,k+1:n) = B(k+1:n,k+1:n)-L(perm(k+1:n),k)·B(k,k+1:n);
         B(k+1:n,k) = zeros(n-k,1);
         B(k,k+1:n) = zeros(1,n-k);
      end
      k = k+1;
   else
      L(perm(k:n),k:n) = eye(n-k+1);
      k = n+1;
   end
end
```

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1993	Technical Report

4. TITLE AND SUBTITLE

Computing Modified Newton Directions Using A Partial Cholesky Factorization

5. FUNDING NUMBERS

DE-FG03-92ER25117
N00014-90-J-1242

6. AUTHOR(S)

Anders Forsgren, Philip E. Gill and Walter Murray

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Operations Research - SOL
Stanford University
Stanford, CA 94305-4022

8. PERFORMING ORGANIZATION REPORT NUMBER

111-MA

9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research - Department of the Navy
800 N. Quincy Street
Arlington, VA 22217

Office of Energy Research
U.S. Department of Energy
Washington, DC 20585

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

SOL 93-1

12a. DISTRIBUTION AVAILABILITY STATEMENT

UNLIMITED

12b. DISTRIBUTION CODE

UL

13. ABSTRACT (Maximum 200 words)

Abstract

The effectiveness of Newton's method for finding an unconstrained minimizer of a strictly convex twice continuously differentiable function has prompted the proposal of various *modified* Newton methods for the nonconvex case.

Linesearch modified Newton methods utilize a linear combination of a descent direction and a direction of negative curvature. If these directions are *sufficient* in a certain sense, and a suitable linesearch is used, the resulting method will generate limit points that satisfy the second-order necessary conditions for optimality.

We propose an efficient method for computing a descent direction and a direction of negative curvature that is based on a partial Cholesky factorization of the Hessian. This factorization not only gives theoretically satisfactory directions, but also requires only a *partial* pivoting strategy, i.e., the equivalent of only two rows of the Schur complement need be examined at each step.

14. SUBJECT TERMS

Unconstrained minimization, modified Newton method, descent direction, negative curvature, Cholesky factorization

15. NUMBER OF PAGES

14 pp

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED			